

UC San Diego

UC San Diego Previously Published Works

Title

Sifting texts: What can a digital tool tell us about late imperial literacy?

Permalink

<https://escholarship.org/uc/item/3jr3z5rt>

Author

SCHNEEWIND, Sarah

Publication Date

2022-12-22

Peer reviewed

Sifting texts: What can a digital tool tell us about late imperial literacy?

© Sarah Schneewind, March 7, 2020.

This talk was prepared for the AAS 2020 in Boston. Please cite it properly if you use it.

Introduction

The Late Imperial Primer Literacy Sieve was designed, by Joshua Day with assistance from historian Jenny Huangfu Day, to answer one research question that I had as I worked on shrines to living men in Ming times. But as it turned out, it showed some other things, too. Other scholars have used it for other research questions. This paper will suggest some more possible uses, although I myself do not plan to pursue them.

I was studying a particular genre of text: essays commemorating shrines to living men. I mostly found the commemorative records in local gazetteers and in Ming men's collected works, but they were also engraved on stones and set up at the shrines. I thought that they sent a particular message. The question I wanted the Late Imperial Primer Literacy Sieve to answer was whether this message would come across to people who could not read the full record, because they did not have a full classical education. So the research question was, "What would a partially-literate Ming person understand from a stone inscription commemorating a premortem shrine?"

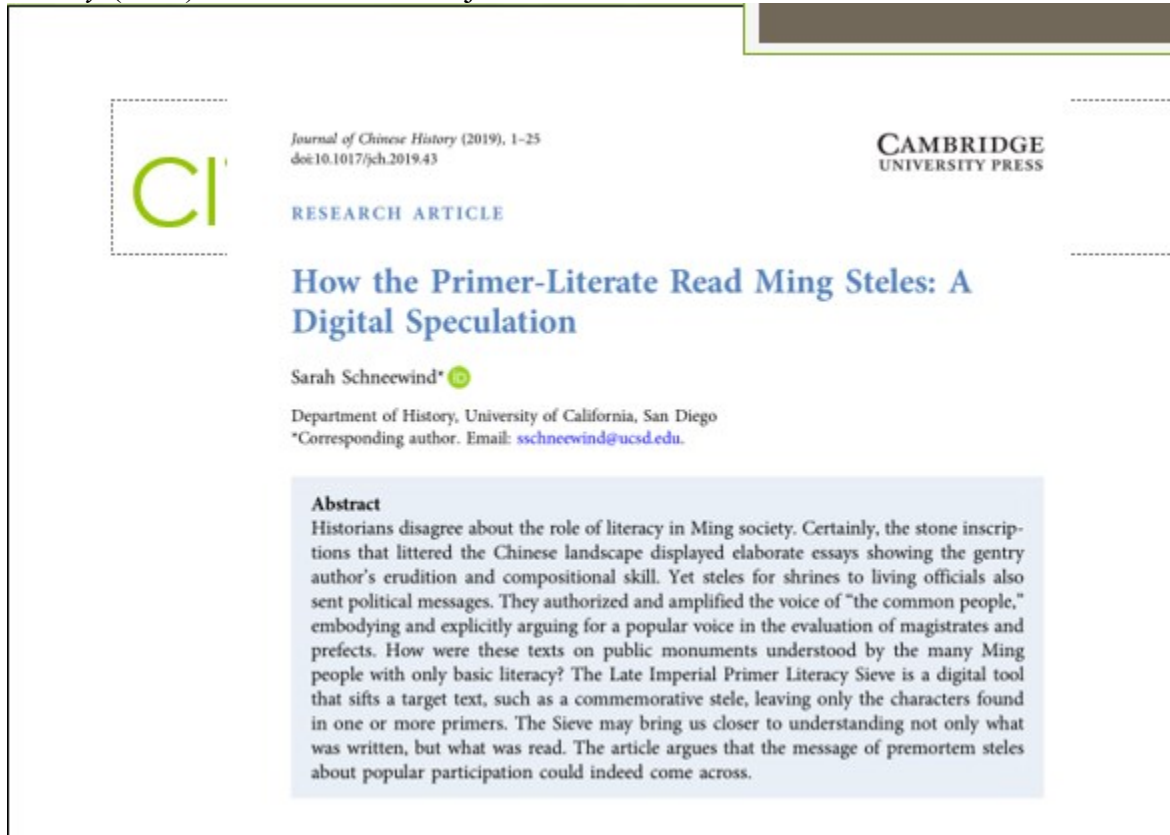
Most people in Ming times probably started their education, and many ended it, with one of a few standard primers. So, the idea of the Sieve was simply to run one or more of the standard primers against the texts I was interested in. It would sift out any characters in the stele text that did not occur in the primer, and I would read the result. That was the basic idea. The designer, Joshua Day, designed it with three elements: primers, target texts, and depleted texts.

The Late Imperial Primer Literacy Sieve : Three Elements

1. Primers

- Built into the Sieve
- Supposedly standard primers for Ming & Qing
- 1 or more selected by the researcher

1. Standard Ming-Qing primers were built into the Sieve, as explained in my recent article, “How the Primer-Literate Read Ming Steles: A Digital Speculation,” *Journal of Chinese History* (2019): 1-25. doi:10.1017/jch.2019.43



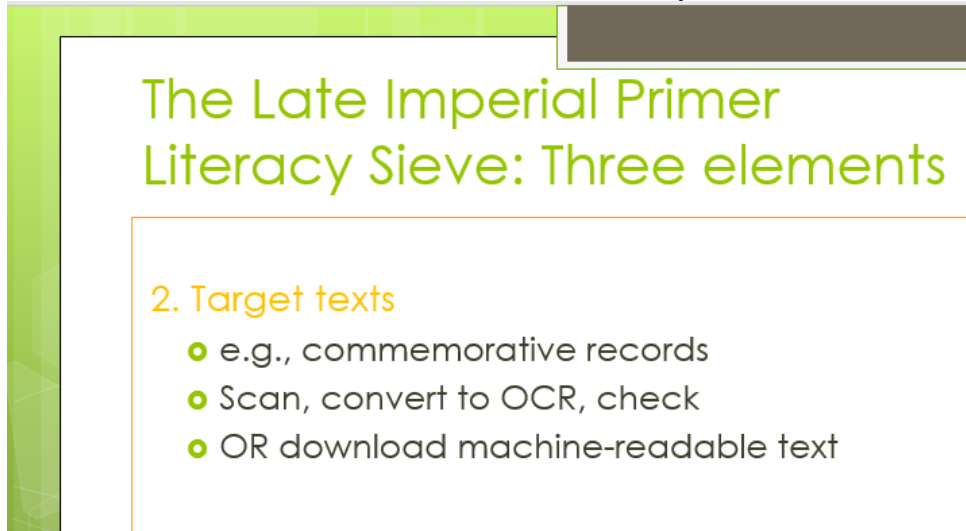
In the lead-up to the panel, Monica Klasing Chen asked whether I had read Wu Huifang 吳蕙芳's work and thought about using 雜字 booklets as primers. Wu considers them to be more 'popular,' and the 千字文, baijiaxing, etc. to have been mainly used by the elite, or at a later stage in education. This would be something that someone could pursue. New primers can be built into the Sieve. The scholar would have to choose which Zazishu was most widespread or applicable to her research.

Should primers include 雜字書 ?

- Wu Huifang 吳蕙芳 thinks them "more popular," the other primers more elite
- Researcher can add any primer
- Few mentions in Ming-era gazetteers

It's actually quite hard to prove what books were used in elementary education; I know because I tried. But even the teachers Li Yu discusses in Qing who were innovating in pedagogy still relied mainly on these standard primers.¹ I noticed that Wu's claim that these books were widely used rests in part on the many editions she found, which is good evidence, but in part on local gazetteers from the 1990s.² When you search the Ming gazetteers in the Erudite data base only one of them mentions one of the texts that Wu talked about.

2. The second element of the Primer Literacy Sieve is the “target texts” -- those I was interested in, or those another research wants to study.

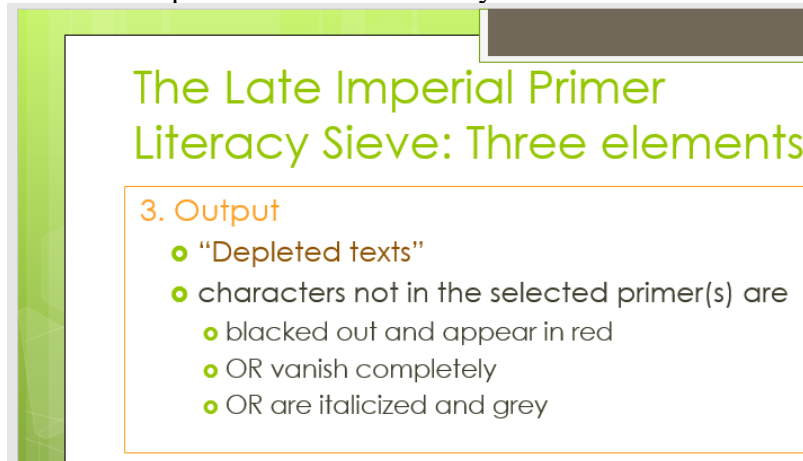


The Late Imperial Primer Literacy Sieve: Three elements

2. Target texts

- e.g., commemorative records
- Scan, convert to OCR, check
- OR download machine-readable text

3. The third element is the output, called “depleted texts”: target texts with the characters not found in primers removed entirely or marked in various ways.



The Late Imperial Primer Literacy Sieve: Three elements

3. Output

- “Depleted texts”
- characters not in the selected primer(s) are
 - blacked out and appear in red
 - OR vanish completely
 - OR are italicized and grey

For a demonstration or to try out the Sieve, please see

¹ Li Yu, “Character Recognition: A new Method of Learning to Read in Late Imperial China,” *Late Imperial China* 33.2 (2012): 1-39.

² 吳蕙芳，明清以來民間生活知識的建構與傳遞 Taibei Shi : Taiwan xue sheng shu ju, 2007, pp. 213 ff for list of gazetteer mentions.

Where is the Sieve?

- knit.ucsd.edu/minghistoryinenglish/late-imperial-primer-literacy-sieve/
- Demonstration:
- paleoludic.com/sieve/ming-literacy-demo.html
- "The Sieve Online," dh.chinese-empires.eu/markus/beta/sieveOnline.html

But this is one way depleted text can appear: the red characters on a black background are not found in the selected primer.

Depleted text

君子之道有三正志定略仁及眾也書曰難大考翼不可征定亂難也而況倉卒之變乎
維忠以正志耳又曰倚乃身迂乃心安民難也而況遠邇一致乎維智以定略耳又曰式
敷民德永肩一心民懷難也而況久思乎維仁以及眾耳三者有一焉皆足以樹勳建名
而況兼之乎此遼之人所以生祀曾公也公代巡茲土遼陽戍卒以撫臣革之貞厲窮兇
者蠟集為亂登陴閉門澳無統紀公時自南衛聞變單騎入城諭之禍福眾心以定蓋示

There are all kinds of problems with “reading” the depleted text, but the *Journal of Chinese History* has just published my discussion of that so I’m not going to talk about that today. Today I’m going to discuss some things that others might want to do with the Sieve.

In designing the Sieve, Josh had three priorities. He followed the guidelines developed by Peter H. Salus.

The Designer's Philosophy

1. Write programs that do one thing and do it well.
2. Write programs to work together.
3. Write programs to handle text [...] because that is a universal interface.

Peter H. Salus
A Quarter-Century of UNIX (1994)

The principles are that programs should each do one thing well, but be able to work with other programs, and be able to handle text. In the case of the Sieve, what this meant to Josh was that: First, the software had to be fast enough to handle a lot of texts. Second, the results had to be something that a historian could interpret – not just statistics but something texty. Third, the results had to be explorable. That is, since it was not clear at the start what results would be most useful, the historian had to be able to use the software to make discoveries.

I got my question answered, but Josh's design meant that other things turned up, too. For instance, it turned out that the amount of one of the target texts that was legible was pretty consistent across the 17 target texts we ran.

How much of a premortem stele could one read?

Roughly, if one had memorized

Heart Sutra: about one-tenth

Hundred Surnames: about one-fifth.

Guanyin Sutra: just under one-third.

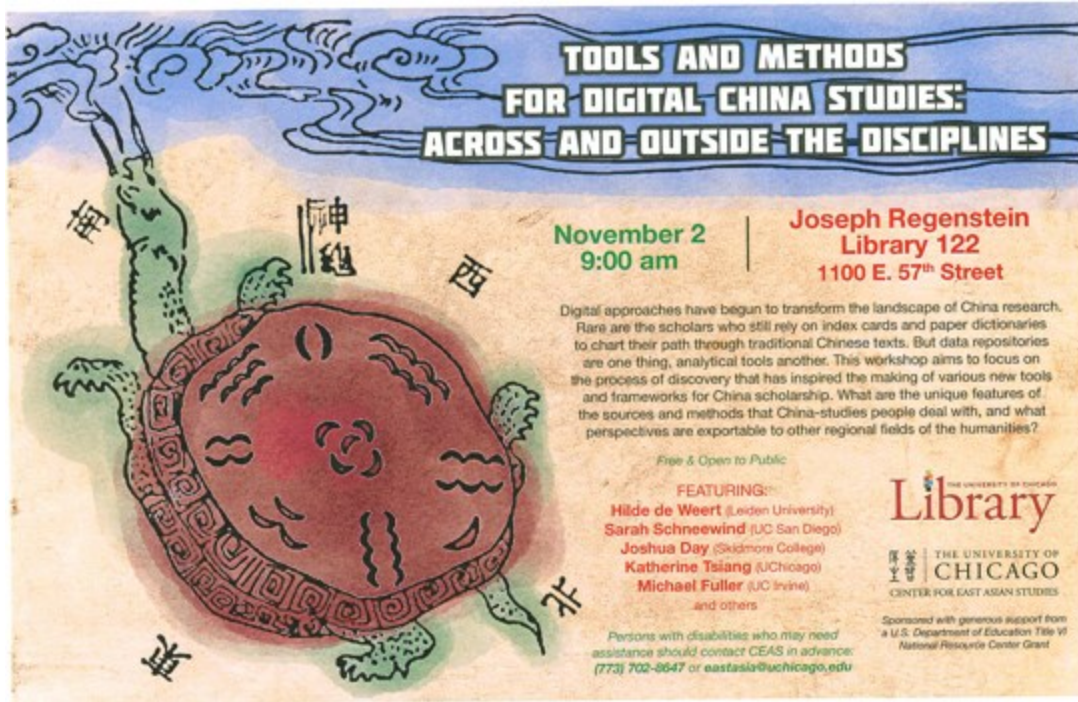
Classic of Filial Piety: just over one-third.

Trimetrical Classic: about 40%.

Thousand Character Essay: a bit over half.

Elementary Learning: 80% or more.

This flexibility means that we can ask other questions. When I went to a workshop at the University of Chicago run by Haun Saussy and Jeff Tharsen recently, the audience did raise other, interesting questions.



Other Questions the Sieve Might Answer

Are the levels of legibility, and the consistency of those levels, a matter of genre? Aside from the two sutras, the other primers have a fairly Confucian outlook, and so do the premortem steles. If the target texts were short stories or Daoist prayers, or legal briefs, or a chunk of the Ming Code, would the levels of legibility be different, and would they vary more? You could run the Guanyin sutra against another sutra, or for instance, the biography of a Buddhist nun, and see if you get better than the 30% it makes legible of premortem steles.

Another question: How much did the primers overlap? That is, if you had learned one, how many words in another would you already know?

This question was raised at the Chicago workshop by Arnd Helmut Hafner, and programmer Leonora Tindall answered it in the Sieve within an hour as the workshop proceeded:

How much did primers overlap?

	三字經	百家姓	般若波羅蜜多心經	孝經	千字文	大學	觀世音普門品經	
三字經	100	25	10	36	58	58	32	
百家姓	26	100	3	16	39	30	18	
般若波羅蜜多心經	46	13	100	43	61	56	65	
孝經	50	22	13	100	67	65	41	
千字文	30	19	7	25	100	42	25	
大學	39	19	8	31	54	100	30	
觀世音普門品經	35	19	16	33	54	50	100	
數字方向	100	26	26	66	60	80	73	100

If you knew→ You could read this percentage of↓	Hundred Surnames	Classic of Filial Piety	Trimetrical Classic	Great Learning	Thousand Character Essay
Hundred Surnames	100	16	26	30	39
Classic of Filial Piety	22	100	50	65	67
Trimetrical Classic	25	36	100	58	58
Great Learning	19	31	39	100	54
Thousand Character Essay	19	25	30	42	100
average	21	27	36	49	55

If you knew that Xiaojing, you could read between 16 and 36% of the other primers, and the most useful one for reading the others would be the Qianziwen, which would let you read over half of each of the others.

Another way to ask this question would be to reverse it. How many more characters would you learn by learning a second primer? That would give us some sense of what families were investing in when they kept a kid in school.

Another kind of question:

Apply this knowledge:

- If a primer was more common in a certain area or time,
- estimate local legibility of texts
- Or estimate temporal legibility of texts

If you knew that a certain primer was more commonly used in a certain place or time – which I must say I think would be very difficult indeed to know – you could estimate how legible other texts would be in that particular social context.

A more interesting question is whether the elite authors of public texts purposely and effectively made them more legible by sticking to the vocabulary of the primers. For a given

author, you could run some elite-only texts, such as his letters to family members or friends, or social poetry, and get a legibility percentage for those. Then you could run something from another genre, such as a public placard or commemorative record, and see if the legibility percentage was higher. Suppose you wanted to know whether Ming officials hoped, when writing certain kinds of memorials, to reach a wider audience? This kind of comparison would be one way to gauge that.

Dan Knorr asked: did officials posted to border areas or poorer, less-educated areas write commemorative essays or other public texts using a smaller vocabulary?

More questions

- ▶ Did the elite write more simply when they were posted to “backward” areas?
- ▶ Did writers design some texts for both the primer literate and the elite?
- ▶ Did vocabulary change within one genre?

Another possible question: how did vocabulary change within one genre over time? For instance, there were premortem shrines in Tang. Was their vocabulary more high-falutin, were they less legible, than in the very different society of Ming?

Improving the Sieve

Hilde de Weerd has already improved the Sieve by making an on-line version. Other developments are certainly possible. For instance:

The Future

- ? Use publicly available datasets to figure out “how readable” particular unknown characters are, on the basis of their constituent radicals.
- ? Take account of more research on what texts teachers used in particular place times.
- ? Run lots of Ming texts to see how many would have been readable with primer literacy

Some people have asked about whether readers could use grammatical predictability to increase bottom-up comprehension. I did not consider this in my original analysis, but there are some tools that might be combined with the Sieve to make that possible to see³:

<https://nlp.stanford.edu/projects/chinese-nlp.shtml> The Stanford Natural Language Processing Group's Chinese Natural Language Processing and Speech Processing

<http://godel.iis.sinica.edu.tw/CKIP/engversion/onlineSystem.htm> KIP Chinese Knowledge and Information Processing

Even without this grammatical sophistication, others have used the tool for other things. Paola Zampierini uses it in her literature classes so students can see how legible vernacular fiction was in Ming times, and one other scholar's use appears in my article. I'll conclude with one more way in which I think the tool *could* be used.

One More Use for the Sieve

Angela Ki Che Leung has written about "Medical Instruction and Popularization in Ming-Qing China." She studies a number of Ming-Qing medical primers. These medical primers refer to medical classics, but use simpler language than those classics and claimed to be for beginners, but adult beginners.⁴ The Sieve offers a way to gauge whether students who had studied a regular primer would do have an easier time learning the medical primers.

But Leung also argues that the language of medical primers became increasingly simple over time, so that those of the late Qing are easier to understand (136). The Sieve, if appropriately modified, could systematize that increasing easiness: one could run the medical primers against the more scholarly works they were based on, and test whether the vocabulary required diminished.

Third, Leung discusses the popularization of medical knowledge in encyclopedias and other sources; the Sieve, again, by running medical primers or classics against a popular medical manual, could pinpoint differences in vocabulary and thus offer a more systematic overview of whether knowledge for professionals differed from that designed for popular use. How many bodies of medical knowledge were there? The Sieve offers one way to begin answering that question, and this method could extend to other fields to suggest how elite and popular knowledge converged or diverged.

I hope that many of you will pick up the tool and adapt it to your own needs. If you find it useful, please do let me know, and please cite it in your work. Many thanks to Monica Klasing Chen for including me in the panel.

³ Chris Atwood commented on these: "My sense is that classical Chinese is so genre- and context-specific that if you wanted machine translation, you'd have to have several versions: Buddhist Classical Chinese, medical classical Chinese, philosophical classical Chinese, historical classical Chinese, and yes, exam essay classical Chinese." Facebook, Sinologists Nov 7 2019. Paul Vierthaler's work on genre may be relevant.

⁴ Angela Ki Che Leung, "Medical Instruction and Popularization in Ming-Qing China," *Late Imperial China* 24.1 (2003): 130-152., 131, 145.

Where is the Sieve?

- knit.ucsd.edu/minghistoryinenglish/late-imperial-primer-literacy-sieve/
- Demonstration:
- paleoludic.com/sieve/ming-literacy-demo.html
- “The Sieve Online,” dh.chinese-empires.eu/markus/beta/sieveOnline.html